Draft-Draft-Draft-Draft-Draft-Draft-Draft-Draft

DRAFT PCI-SIG ENGINEERING CHANGE NOTICE

TITLE:	Address Translation Relaxed Ordering
DATE:	May 17th, 2018
AFFECTED DOCUMENT:	PCI_Express_Base_4.0r1.0 Specification
SPONSOR:	Eric Wehage Huawei Technologies Co.,Ltd.

Part I

1. Summary of the Functional Changes

This ECN allows Address Translation Requests and Completions to support the Relaxed Ordering bit, where are currently defined to be Reserved for these types of TLPs.

The proposal preserves interoperability with older Translation Agents.

2. Benefits as a Result of the Changes

Address Translation Requests are already allowed to bypass any type of traffic to the TA, however, the Address Translation Completions are not allowed to bypass other Completions or Posted requests when returning to the ATC, even though these Address Translations have no relationship with other traffic.

By allowing the Address Translation Completions to bypass other traffic using the Relaxed Ordering bit, it can allow Root Ports, Switches, and Endpoints to prioritize the Address Translation Completions in the same manner that Relaxed Ordered completions are, potentially improving the return latency for these critical translations.

3. Assessment of the Impact

The requested change is fully backward compatible with older TA and ATC hardware.

Interoperability with TA's conforming to older versions of the ATS specification is retained. They would continue to return Translation Completions with RO bit clear. The ATC would be required to accept the Translation Completion even if the RO bit didn't match the Translation Request RO bit.

All future TA's must ensure that the RO bit is copied from the Translation Request to the Translation Completion (note that this is already required by older versions of the specification, but some TA's may not have followed the rule since the Attr fields were listed as reserved).

ECRC checks are not impacted, since ECRC is already calculated using RO (even if RO is reserved for some packet types).

Translation Completions are not required to be ordered with Invalidation Requests (this is already explicitly stated in the ATS specification).

Older ATC hardware would not be impacted when sending requests to a newer TA hardware, since the older ATC hardware would only send requests with RO bit clear and the newer TA hardware would copy the RO as-is to the completion.

Older switches would not be impacted, since they are already required to pass the Attr field through in all requests and completions, including Translation Requests and Translation Completions.

4. Analysis of the Hardware Implications

Translation Agents would be required to copy the RO from the Translation Request to the Translation Completion. This is already required by the PCIe Specification in section 2.2.9.

Address Translation Caches would be optionally allowed to generate Address Translation Requests with RO bit set.

No new spec defined control bits would be required. The existing Enable Relaxed Ordering bit in the Device Control register would apply to Translation Requests as they do for all requests. A new capability bit is added to the ATS Capability Structure to advertise this capability. Proprietary control may be implemented to enable RO in the Translation Requests by the ATC.

5. Analysis of the Software Implications

There would be no additional software requirements. ATC driver or TA driver software may choose to enable or disable use of RO in a proprietary manner.

6. Analysis of the C&I Test Implications

This ECN allocates previously reserved bits. A few new C & I tests would be required to cover the use of RO for Translation Requests and Completions. Existing Translation tests would need to be modified to allow RO to be set.

Part II

Detailed Description of the change

2 Transaction Layer Specification

:

:

2.2Transaction Layer Protocol - Packet Definition

:

:

2.2.9 Completion Rules

:

:

- Completion headers must supply the same values for the Attribute as were supplied in the header of the corresponding Request, except as explicitly allowed when IDO is used (see Section 2.2.6.4).

:

:

10 ATS Specification

10.2 ATS Translation Services

10.2.2 Translation Requests

A Translation Request has a format that is similar to that of a memory read. The AT field is used to differentiate a Translation Request from a normal memory read.

The request header for a Translation Request has the formats illustrated in Figure 10-8 and Figure 10-9.

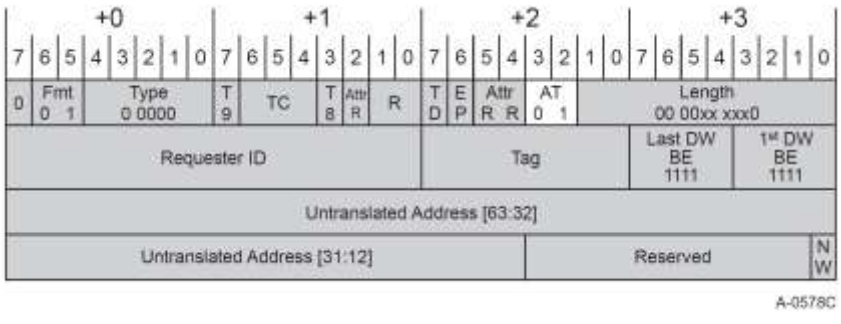


Figure 10-8: 64-bit Translation Request Header

Commented [EW1]: The Attr field needs to indicate that RO bit is not reserved.

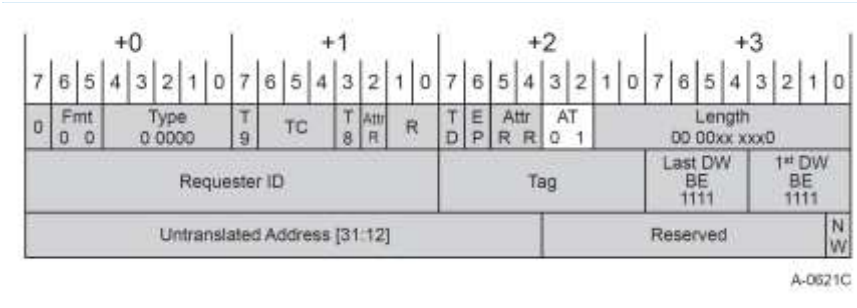


Figure 10-9: 32-bit Translation Request Header

Commented [EW2]: The Attr field needs to indicate that RO bit is not reserved.

Translation Requests have the same completion timeout intervals as Read Requests.

10.2.2.1 *Attribute Field*

For a Translation Request, the Relaxed Ordering (RO) bit may be set if Enable Relaxed Ordering in the Device Control Register is set on an implementation specific basis. The remainder of the Attr field is reserved for future use.

–There are no ordering requirements for a Translation Request. A TA may reorder a Translation Request with respect to any other request.

IMPLEMENTATION NOTE

Translation Request Ordering

Because no ordering can be assumed between Translation Requests and other types of Requests, a Translation Request does not make an effective flushing/ordering primitive.

:

:

:

10.2.3 Translation Completion

A Translation Completion (either a Cpl or a CplD) is sent by a TA for each Translation Request. This specification describes the meaning of fields in Translation Completions. Fields not defined in this specification have the same meanings proscribed for Read Completions in this specification. The Attr field is reserved for future use.

If the TA was not able to perform the requested translation, a completion with the format shown in Figure 10-10 is used.

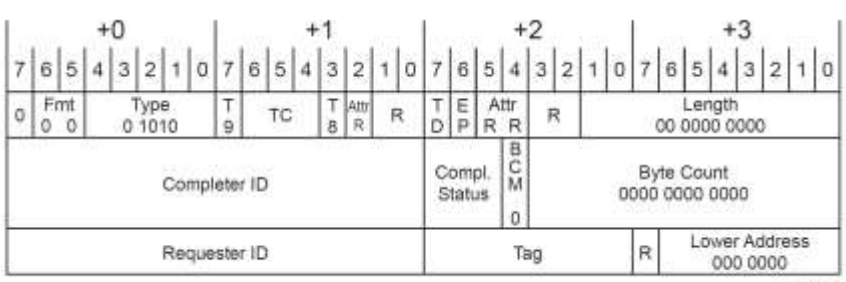
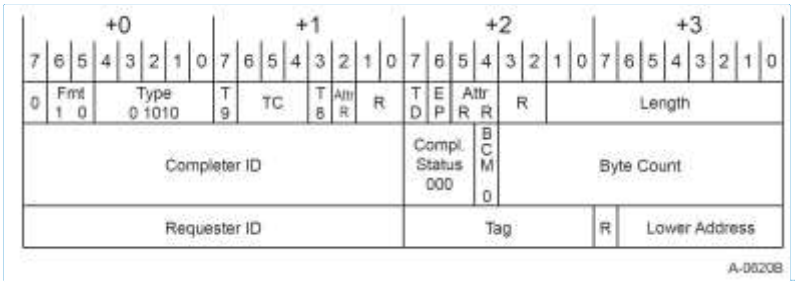


Figure 10-10: Translation Completion with No Data

:

:

:



Commented [EW3]: The Attr field needs to indicate that RO bit is not reserved.

Commented [EW4]: The Attr field needs to indicate that RO bit is not reserved.

Figure 10-11: Successful Translation Completion

Fields are set in accordance with Sections 2.2.9 and Section 2.3.1.

Translation Completions must be sent using the same TC as the Translation Request. The Function is not required to verify that the same TC was used.

A TA may optionally copy the RO bit of a Translation Request to the Translation Completion in accordance to the rule specified for the attribute field of completions in Section 2.2.9. If a TA does not copy the RO bit of a Translation Request to the Translation Completion, the TA must clear the RO bit in the Translation Completion. However, it is strongly recommended that the TA copy the RO bit.

A Translation Completion with RO set may follow the ordering rules for Relaxed Ordered Completions as specified in Section 2.4.1.

The function that initiated the Translation Request must not report an error if the Translation Completion RO bit was clear when the corresponding Translation Request RO bit was set.

IMPLEMENTATION NOTE

Attribute Field Compatibility in Translation Completions

Some implementations of TA may not copy the Attribute field from the request to the completion as required by Section 2.2.9, since the Attr field is defined as reserved in previous versions of the ATS Specification.

Therefore, the following situations may occur and are handled as follows:

- A TA that does not copy the RO bit (as is typically done for completions as indicated in Section 2.2.9) by forcing RO to 0 is coupled with a function conforming to this specification that allows RO to be set in the request. The function will accept a Translation Completion with RO clear and not log an error.
- A TA that conforms to the Attr copy rule (in Section 2.2.9) is coupled with a function that does not support RO in Translation Requests. Translation Completions will return with RO clear as the function expects.

Therefore, the use of RO is made fully backward compatible. However, it is strongly recommended that the TA support the copy of the RO bit conforming to the rules in Section 2.2.9.

Commented [EW5]: I wish rules were numbered in each section. Specific cross reference would help.

The Lower Address field will contain a value that will make the packet consistent with RCB semantics. If the result is returned in a single packet, Lower Address is set to RCB minus Byte Count. If the results are returned in multiple packets, the first packet will have a Lower Address field of RCB minus (Total Completion Length * 4) and subsequent packets will have a Lower Address field of 000 0000b. See Section 10.2.4 for additional requirements for multiple packet completions.

If the Completion Status field is 000b, then the translation was successful and a data payload will follow the header. The contents of the data payload are shown in Figure 10-12.

10.2.3.1 *ATS Capability Register*

Figure 10-21 details the allocation of register fields of an ATS Capability register; Table 10-9 provides the respective bit definitions.

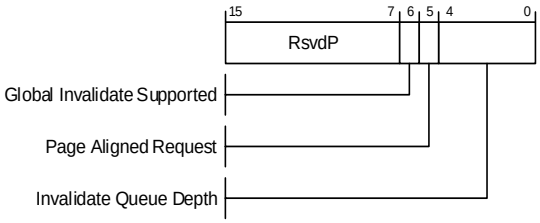


Figure 10-21: ATS Capability Register

Table 10-19: ATS Capability Register

Bit Location	Register Description	Attributes
4:0	**Invalidate Queue Depth** – The number of Invalidate Requests that the Function can accept before putting backpressure on the Upstream connection. If 0 0000b, the Function can accept 32 Invalidate Requests.	RO
5	**Page Aligned Request** – If Set, indicates the Untranslated Address is always aligned to a 4096 byte boundary. Setting this field is recommended. This field permits software to distinguish between implementations compatible with this specification and those compatible with an earlier version of this specification in which a Requester was permitted to supply anything in bits [11:2].	RO
6	**Global Invalidate Supported** – If Set, the Function supports Invalidation Requests that have the Global Invalidate bit Set. If Clear, the Function ignores the Global Invalidate bit in all Invalidate Requests (see Section 10.3.8). This bit is 0b if the Function does not support the PASID TLP Prefix.	RO
7	**Relaxed Ordering Supported** – If Set, indicates this function may set the RO bit in Translation Requests when Enable Relaxed Ordering bit is set in the Device Control Register of the PCI Express Capability Structure. Also indicates that Translation Completions RO bit clear will not be considered an error when the Translation Request RO bit was set.	RO